The *Perfect* Fit

Like **Cinderella's glass slipper**, the interaction between many biomolecules relies on a perfect fit. Humans have taken advantage of this recognition-binding behavior in research and medicine for over a century.

IN A HOME PREGNANCY TEST, something inside the test stick reacts with something in the sample to provide a yes or no answer. The stakes are high: if the something inside the test doesn't properly identify its target, there could be a lot of unhappy people.

A pregnancy test is an example of an immunoassay—a test used in molecular biology labs worldwide. Immuno-assays take advantage of immune-system proteins called antibodies that recognize and bind to specific target molecules, confirming their presence in a sample. This general idea of molecular recognition is a powerful concept used by living organisms for nearly all cellular processes. It is well known that antibodies recognize pathogens, but cells also use recognition proteins for many other functions, such as on-off switches for gene transcription or enzyme production, and as receptors for communication between cells. In addition, many drugs and therapies work by targeting at least one partner in these binding relationships. For example, some drugs work by blocking receptors to inhibit pathogen growth.

Antibodies used for clinical purposes—as well as for home pregnancy tests—are highly regulated to ensure they detect what they claim to detect. However, Los Alamos molecular biologist Andrew Bradbury is concerned that the antibodies used in general laboratory experiments are not. This is a problem not only because it hampers scientific discovery, but also because these basic science experiments are foundational to the development of new drugs that might someday make it to market.

A 2008 study revealed that less than 50 percent of the 6000 routinely used commercial antibodies properly recognized their targets. Bradbury is adamant that this is a major problem. In fact, he and 111 of his colleagues recently published a commentary stating their concern that nearly $350 million is wasted annually, in the United States alone, on antibodies that are nonfunctional.

"This $350 million does not take into account the money wasted on scientists' time, other reagents, or attempts to validate research that is not reliable or reproducible," says Bradbury.

The good news is that a solution is known: More reliable antibodies can be generated in the laboratory based on the genetic code that defines them. In fact, Bradbury's team at Los Alamos has already created a successful pipeline for making highly specific antibodies of this kind in a high-throughput, standardized fashion. And another team at Los Alamos has developed a way to create other types of binding proteins using specialized algorithms to customize their shape for any desired target. Together, the researchers' hope is that these technologies, alongside increased demand from the scientific community, will change the paradigm for using standardized molecular recognition in research once and for all.

Building without a blueprint

Antibodies are the little red flags of the immune system. They are Y-shaped proteins that recognize and bind to antigens, foreign substances in the body, tagging them for destruction by other immune system cells. Antibodies and their immune system cohorts, memory B cells, work together to remember foreign invaders in order to respond more quickly the next time they are encountered. This is the basis for vaccination: inducing immunity by introducing a dead or inactivated strain of a pathogen so a person can create antibodies against it and B cells to remember it.

Capitalizing on this principle, antibodies for use in research and medicine are made by inoculating animals (usually mice and rabbits) with the desired antigen, such as a virus, and isolating the resulting antibodies. The problem with this method is that the antibodies aren't identical—they're polyclonal, meaning they're all slightly different from each other but still bind to the same antigen—and are mixed with other antibodies for other targets. In fact, only about 0.5–5 percent of the antibodies extracted from an animal host are sufficiently specific to be useful in the lab for reliable recognition.
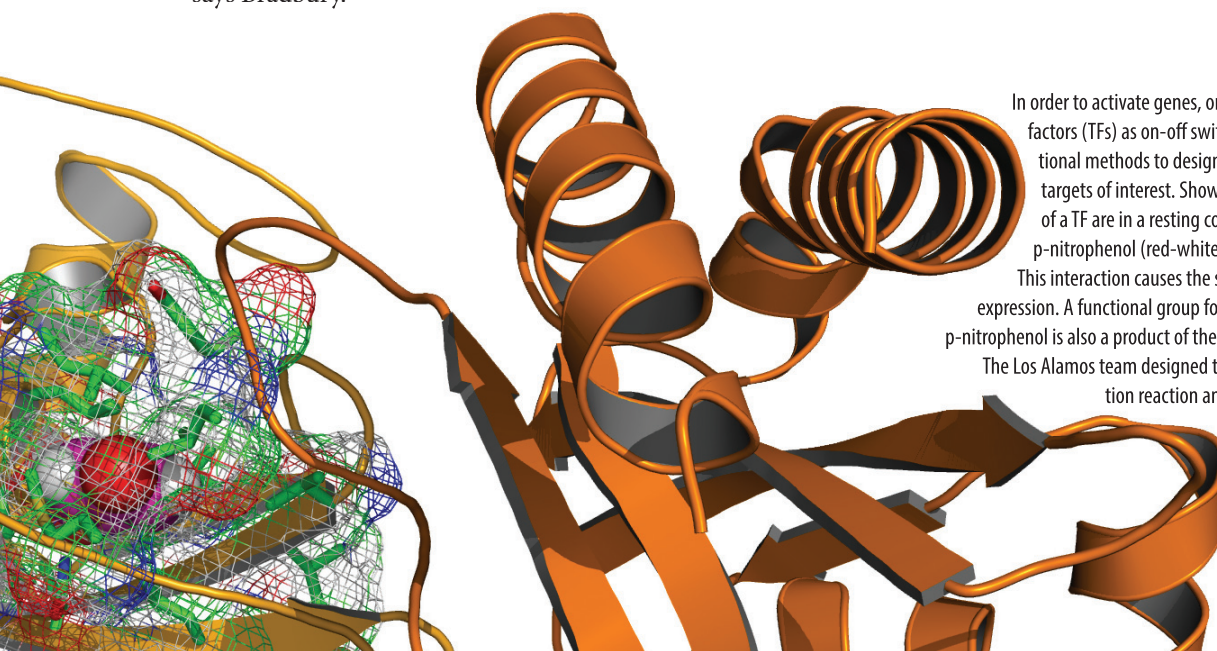
In order to activate genes, organisms use molecules called transcription factors (TFs) as on-off switches. Los Alamos scientists use computational methods to design novel TFs uniquely matched to specific targets of interest. Shown here, the two domains (gold and orange) of a TF are in a resting conformation until the target molecule, p-nitrophenol (red-white structure), enters its binding site (mesh). This interaction causes the shape of the TF to adjust and activate gene expression. A functional group found in thousands of commercial chemicals, p-nitrophenol is also a product of the degradation of the insecticide paraoxon. The Los Alamos team designed the TF as a biosensor to detect this degradation reaction and also designed the enzyme that degrades paraoxon in the first place—all based on the idea of molecular recognition.

Nileena Velappan prepares phage and yeast cells to screen for high-affinity antibodies. The pipeline her team uses ensures the reliability and reproducibility of antibodies for use in research by selecting them based on their genetic blueprint.

An improvement to this method came about in the 1970s, when scientists created "hybridomas" in the lab by taking antibody-producing B cells from inoculated mice and fusing them with cancer cells. In theory, the cancer cells enable the hybrid to grow indefinitely, creating an endless supply of antibodies—all produced from the same B cell, so that the resulting antibodies would be monoclonal (identical). Unfortunately, these hybrid cell lines are not as great as they sound. They do not actually last forever, endlessly producing antibodies, and their antibodies require careful characterization because some of the antibodies produced still recognize more than one target.

The problem is that both of these approaches still rely on design based on prior exposure, rather than based on specific instructions. It's somewhat akin to an architect recreating the Eiffel Tower based on his or her memory of seeing it at age 10, instead of using a blueprint. It would be more accurate to use the blueprint.

"Our work creates antibodies that are defined by their DNA sequences, just as genes are," says Bradbury. "We believe the entire community should move to methods of generating antibodies that will not require the use of animals at all. This will directly lead to more reliable molecules."

Using a blueprint library

In order to create an antibody with a known genetic blueprint, a scientist needs two things: the antibody's genetic sequence and the molecular machinery to translate it into protein. The molecular machinery part can easily be found inside a bacterium or other cell; however, to obtain an antibody's genetic sequence one must first know which antibody perfectly fits the antigen of interest.

For this process, Nileena Velappan and Leslie Naranjo, on Bradbury's Los Alamos team, begin by using an antibody library Bradbury developed more than 15 years ago. The library was made by taking lymphocytes, which include B cells that produce antibodies, from the blood of 40 donor
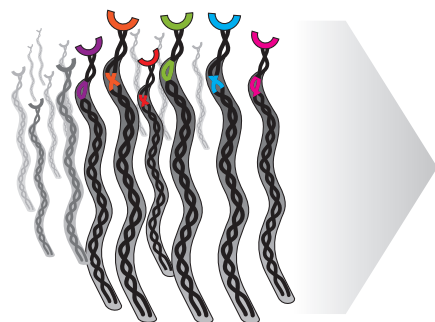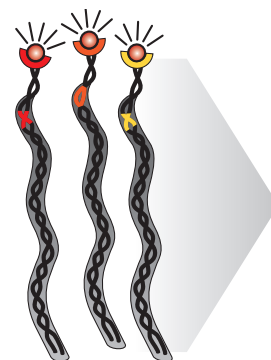
Phage Display

Library of billions of filamentous bacteriophages, each displaying one human antibody (encoded for by the corresponding gene in the phage DNA)
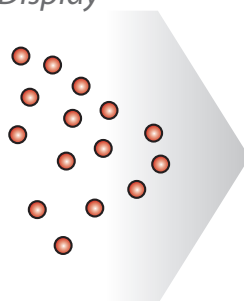
A target, such as ricin, is added to screen the library for matches.

Target binds to antibodies with a good-affinity match.

individuals and then extracting all the segments of DNA that encode various antibodies. These antibodies are unknown at this point; the majority represents a mixture of the antibodies generated by the donor's immune system in the past, some from exposure to a pathogen, either naturally or through vaccination.

The next step in the process is to put the DNA into an organism that can use it to make the antibodies. For this, the team uses a filamentous bacteriophage—a long thin virus that infects bacteria such as *E. coli*. A phage, as it is more commonly called, possesses only a simple loop of single-stranded DNA that encodes its entire genome. Included in this genome is the code for a viral protein coat, which is an outer protective layer, and when the sequence for an antibody is inserted into the protein coat section of the phage genome, the phage—using the molecular machinery of its host *E. coli* bacterium—builds the desired antibody and "displays" it, sticking it on the outside of its protein coat. This is called phage display.

Armed with a library of antibodies, one stuck on the surface of each phage, the team can introduce a target antigen—for example, ricin. The scientist relies on the fact that when the complementary antibody and antigen bind together, the phage carrying the matched antibody can be captured and its antibody gene extracted. This ensures that the gene for the best-matching antibody can be replicated again and again with certainty that it produces the right-fit antibody for ricin. This phage display library technique is especially useful for selection from large libraries with up to 200 billion phage-antibody combinations from which to choose. However, when a match is found, it is still laborious for the scientist to determine the affinity, or quality of the match.

To tackle this aspect, Bradbury's team introduces a second round of selection using yeast cells. The antibody genes from the chosen phage are inserted into the genetic material of yeast—again, one antibody choice per yeast cell—and the cells make thousands of antibodies and display them on their surfaces. The advantage here is that the diversity of potential matches has been reduced (during the first round

in phage) to closer to 1 million antibody choices, and a significant proportion of them have already shown some level of affinity towards the target. Furthermore, yeast cells are much bigger than phage and can be visualized using a cell-sorting machine called a flow cytometer to separate the yeast cells that have a bound antigen.

"Here, we play with changing the concentration of the target antigen until we find the lowest concentration that still binds, indicating a high-affinity match," says Bradbury. Once this perfect fit is found, the antibody gene is extracted from the yeast and sequenced. Together, this phage and yeast display process enables high-throughput antibody selection. Once sequenced, the gene for the chosen antibody can be used to produce the exact desired antibody whenever it is needed.

In 2011, Bradbury's team was awarded a grant from the National Institutes of Health to develop a pipeline capable of creating antibodies for all human proteins. By looking at antibodies, researchers can identify where in a cell genes are active and under what conditions they increase or decrease their expression. In addition to this basic research on cellular functions, antibodies are also useful in the development of therapeutics and vaccines. Many therapeutics exploit the highly specific binding of antibodies to deliver drugs to cells, such as cancer cells. In fact, Bradbury has had a collaboration with researchers at the University of New Mexico since 2013 to create antibodies against known cancer targets.

When the library is empty

The display methods used by Bradbury's team work well for known target molecules that are proteins. But sometimes the body is exposed to new or rare targets that are not proteins and for which there is no library of recognition molecules. Pesticides, industrial chemicals, and even nerve gases are examples of manmade substances to which humans have had no prior exposure (until recently) that might produce adverse side effects. The reason humans and other organisms have not developed recognition methods against these chemicals is not lack of ability but rather lack of need. Library or not, biological molecular recognition still could
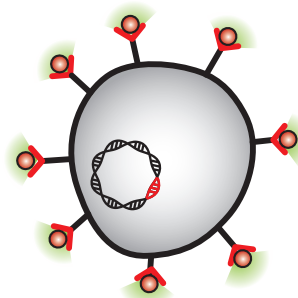
Yeast Display
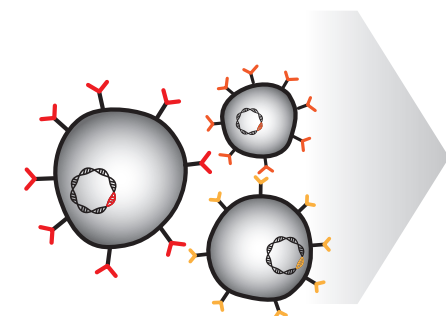
Genes from well-matched phage are added to the genetic material in yeast cells, directing the yeast to display antibodies.

Target ricin is added again, at differing concentrations, and highest-affinity matches are selected using flow cytometry.

The highest-affinity match is selected, indicating that the gene inside this yeast encodes the perfect-fit antibody for ricin.

be the most effective route to the development of sensors to detect toxic chemicals—and even enzymes to neutralize or destroy them.

"Molecular recognition is a major part of what distinguishes biology from chemistry," says Los Alamos biologist Charlie Strauss. One molecule being able to recognize and exclusively interact with another is an essential component of cellular metabolism, communication, reproduction, and evolution. "The beauty of recognition proteins is that this single class of molecule has variants that can grip any target structure with sub-angstrom [less than one ten-billionth of a meter] shape complementarity. Change the gene sequence and you change the shape—it's that simple. No class of manmade materials achieves that degree of programmable molecular recognition, so this paradigm is only seen in biology."

With this in mind, Strauss and colleague Ramesh Jha, also a biologist at Los Alamos, have created an artificial selection system to identify the best binding molecules for specific targets. They do so by combining computation and experimental tools to create recognition-binding proteins from scratch. Without a known library of proteins to choose from, they instead rely on a computer to suggest a number of proteins whose shape will likely bind to the desired target. They use Rosetta protein-modeling software, a commercial platform that Strauss helped originate in 1998, to help manage the possibilities. For a small protein, say just 100 amino acids long, there could be up to 10^{180} possible genes that produce a matching conformation. "That's more than there are atoms in the observable universe by a factor of a googol," says Strauss. "So we use a computer model to winnow it down to a feasible genetic library of a *mere* billion guesses."
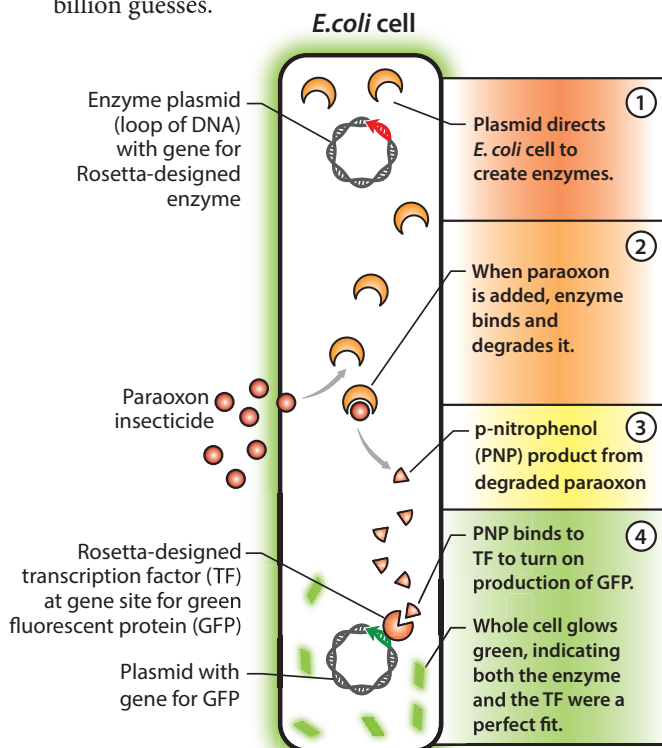
"Even so, manually testing, let alone synthesizing, a billion of anything is prohibitive, so instead we let *E. coli* cells do both for us," says Jha. The trick is to make the binding event itself trigger a gene that makes the cell glow. Each cell carries within it a unique binding candidate from the Rosetta-developed library. If the candidate is able to successfully bind, the cell carrying it expresses a fluorescent protein.

One-stop shopping

"Creating a protein from scratch that fits a target molecule is the ultimate test of our understanding of protein design," says Jha. High-affinity binding is a complicated interaction between molecules, made possible by atoms located in the right spaces to make bonds and interactions. High affinity allows the protein to capture its specific target exclusively, even in the complex mixture of similar chemicals found inside a cell.
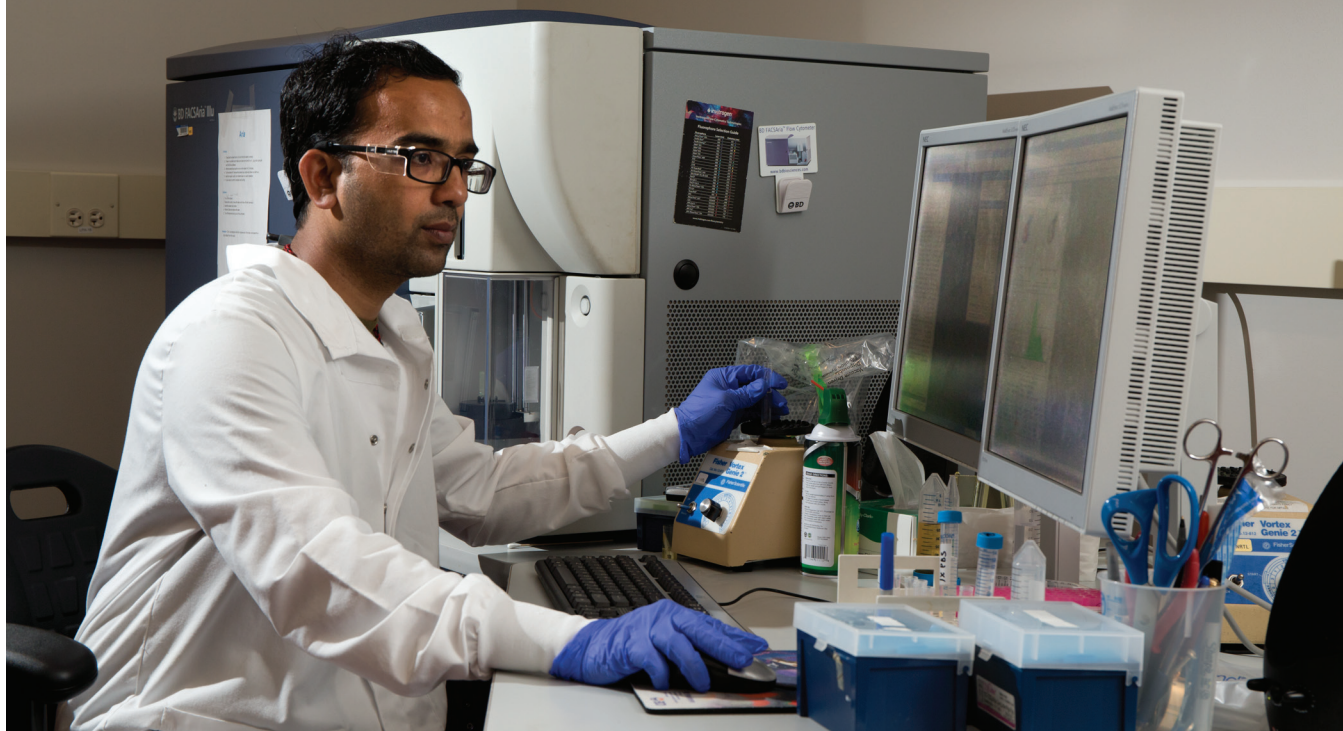
Currently, Strauss and Jha are working on enzymes that catalyze the degradation of paraoxon, which is a powerful insecticide that works by inhibiting a neurotransmitter. This happens to be similar to the way sarin gas and many other nerve agents work, making paraoxon a fairly dangerous chemical that is now very rarely used in agriculture.

The Rosetta software analyzed the shape of the target paraoxon molecule to determine the complementarily shaped binding site that a protein would need to create a high-affinity match. The team used the software to select a known enzyme as a scaffold that could support the genetic adjustments needed to create new chemical activity in the enzyme—in this case, to hypothetically bind and degrade paraoxon. This alone would be a fantastic result: a custom-made enzyme that can degrade a powerful nerve agent.

Once Rosetta suggested billions of enzyme variants to try, the team engineered *E. coli* cells to produce the enzymes (one version of the enzyme per bacterium). However, individually testing the chemical contents of billions of cells for a dilute reaction product is prohibitively difficult, so the team also added a biosensor for the reaction product to the *E. coli* cell.

"When paraoxon is degraded, one of the products is a molecule called p-nitrophenol, or PNP," says Jha. "By engineering a protein biosensor that recognizes the PNP, we can determine that the paraoxon was successfully degraded." For this, the team chose a transcription factor as the scaffold for their second custom binder and used Rosetta to determine the code to make it bind to the PNP product. Transcription factors turn on the expression of genes, so the team added the gene for green fluorescent protein—which causes cells to glow green—to be used as a signal to show that PNP (and paraoxon degradation) has been detected in the cell.

E.coli cell

Enzyme plasmid (loop of DNA) with gene for Rosetta-designed enzyme

① Plasmid directs *E. coli* cell to create enzymes.

② When paraoxon is added, enzyme binds and degrades it.

Paraoxon insecticide

③ p-nitrophenol (PNP) product from degraded paraoxon

Rosetta-designed transcription factor (TF) at gene site for green fluorescent protein (GFP)

④ PNP binds to TF to turn on production of GFP.

Plasmid with gene for GFP

Whole cell glows green, indicating both the enzyme and the TF were a perfect fit.

Ramesh Jha prepares a sample for screening in a flow cytometer. Flow cytometry, which was invented at Los Alamos in the 1960s, is a valuable research tool used for high-throughput cell sorting.

This setup, which couples the catalytic event within each individual cell to a fluorescence signal, makes it possible to separate the cells using a flow cytometer. The glowing cells with a successfully functioning enzyme can be isolated and allowed to reproduce so the sequences for the successful enzymes can be retrieved for further analysis and iteration with Rosetta.

The team now has three results: a customized enzyme that can bind to a chemical and catalyze its modification, a transcription factor protein that can detect the reaction product, and a system for testing the effectiveness of each at the single-cell level. Whew.

Matchmaker, matchmaker

What's significant about the two methods of obtaining molecular recognition proteins—from antibody libraries or by Rosetta design—is that the binding proteins are defined by their genetic codes so they can be reproduced reliably and sustainably. In addition, they both use flow cytometry (which was invented at Los Alamos, as described in the November 2013 issue of *1663*) to make high-throughput screening possible. This dramatically impacts the efficiency of the laboratory processes, making the selection methods both specific and prolific.

Reliability, sustainability, and speed are important characteristics of the next generation of biological research tools. Now that genomic sequencing is faster and readily available to scientists, the next big challenge to understanding living systems is assigning function: matching proteins and enzymes to their genetic blueprint. In the past, this could only be accomplished by a trial-and-error approach, but new technologies, some coupled with computation, are demonstrating entrance to a new era.

This trend will open doors to many possibilities. The most obvious ones are in the medical arena, which is especially important now that microbial resistance is increasingly rendering existing drugs ineffective. Molecules that recognize targets are the basis of many drugs and therapeutics, and being able to find and test the affinity of potential new ones, as well as quickly screen them, will be critical. Furthermore, engineering custom catalysts, as demonstrated in the paraoxon detector, has a multitude of applications, ranging from production of biofuel precursors to fossil-fuel-free nylon to enzymes that help with materials manufacturing. The key to it all is having the right fit.

—Rebecca McDonald

More **protein research** at Los Alamos

Protein structure is key to overall protein function. Los Alamos scientists are investigating ways to analyze these structures using a number of techniques:

Determining protein structures with SOLVE/RESOLVE and Phenix software

Determining protein structures with neutron crystallography

Ribosome modeling

More protein science at Los Alamos